

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/534,130A
Source: PUF
Date Processed by STIC: 5/23/06

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 05/23/2006

PATENT APPLICATION: US/10/534,130A

TIME: 14:02:49

Input Set : A:\21415015.APP

Output Set: N:\CRF4\05232006\J534130A.raw

3 <110> APPLICANT: FRAZER, IAN HECTOR

5 <120> TITLE OF INVENTION: A METHOD FOR OPTIMISING GENE EXPRESSION USING

SYNONYMOUS

6 CODON OPTIMISATION

8 <130> FILE REFERENCE: 21415-0015US

10 <140> CURRENT APPLICATION NUMBER: 10/534,130A

11 <141> CURRENT FILING DATE: 2005-05-06

13 <150> PRIOR APPLICATION NUMBER: PCT/AU03/01487

14 <151> PRIOR FILING DATE: 2003-11-10

16 <150> PRIOR APPLICATION NUMBER: 60/425,163

17 <151> PRIOR FILING DATE: 2002-11-08

19 <160> NUMBER OF SEQ ID NOS: 126

21 <170> SOFTWARE: PatentIn version 3.3

23 <210> SEQ ID NO: 1

24 <211> LENGTH: 714

25 <212> TYPE: DNA

26 <213> ORGANISM: Artificial Sequence

28 <220> FEATURE:

29 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

30 humanized GFP sequence

32 <220> FEATURE:

33 <221> NAME/KEY: CDS

34 <222> LOCATION: (1)..(711)

36 <400> SEQUENCE: 1

37 agc aag ggc gag gaa ctg ttc act ggc gtg gtc cca att ctc gtg gaa 48

38 Ser Lys Gly Glu Leu Phe Thr Gly Val Pro Ile Leu Val Glu

39 1 5 10 15

41 ctg gat ggc gat gtg aat ggg cac aaa ttt tct gtc agc gga gag ggt 96

42 Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly

43 20 25 30

45 gaa ggt gat gcc aca tac gga aag ctc acc ctg aaa ttc atc tgc acc 144

46 Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr

47 35 40 45

49 act gga aag ctc cct gtg cca tgg cca aca ctg gtc act acc ttc tct 192

50 Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser

51 50 55 60

53 tat ggc gtg cag tgc ttt tcc aga tac cca gac cat atg aag cag cat 240

54 Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln His

55 65 70 75 80

57 gac ttt ttc aag agc gcc atg ccc gag ggc tat gtg cag gag aga acc 288

58 Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr

59 85 90 95

61 atc ttt ttc aaa gat gac ggg aac tac aag acc cgc gct gaa gtc aag 336

62 Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/534,130A

DATE: 05/23/2006

TIME: 14:02:49

Input Set : A:\21415015.APP

Output Set: N:\CRF4\05232006\J534130A.raw

```

63          100          105          110
65 ttc gaa ggt gac acc ctg gtg aat aga atc gag ctg aag ggc att gac      384
66 Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp
67          115          120          125
69 ttt aag gag gat gga aac att ctc ggc cac aag ctg gaa tac aac tat      432
70 Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr
71          130          135          140
73 aac tcc cac aat gtg tac atc atg gcc gac aag caa aag aat ggc atc      480
74 Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile
75 145          150          155          160
77 aag gtc aac ttc aag atc aga cac aac att gag gat gga tcc gtg cag      528
78 Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln
79          165          170          175
81 ctg gcc gac cat tat caa cag aac act cca atc ggc gac ggc cct gtg      576
82 Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val
83          180          185          190
85 ctc ctc cca gac aac cat tac ctg tcc acc cag tct gcc ctg tct aaa      624
86 Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys
87          195          200          205
89 gat ccc aac gaa aag aga gac cac atg gtc ctg ctg gag ttt gtg acc      672
90 Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr
91          210          215          220
93 gct gct ggg atc aca cat ggc atg gac gag ctg tac aag tga      714
94 Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
95 225          230          235
98 <210> SEQ ID NO: 2
99 <211> LENGTH: 237
100 <212> TYPE: PRT
101 <213> ORGANISM: Artificial Sequence
103 <220> FEATURE:
104 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
105      humanized GFP sequence
107 <400> SEQUENCE: 2
108 Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu
109 1          5          10          15
112 Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly
113          20          25          30
116 Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr
117          35          40          45
120 Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser
121          50          55          60
124 Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln His
125 65          70          75          80
128 Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr
129          85          90          95
132 Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys
133          100          105          110
136 Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp
137          115          120          125

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/534,130A

DATE: 05/23/2006

TIME: 14:02:49

Input Set : A:\21415015.APP

Output Set: N:\CRF4\05232006\J534130A.raw

```

140 Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr
141      130                      135                      140
144 Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile
145 145                      150                      155                      160
148 Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln
149                      165                      170                      175
152 Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val
153                      180                      185                      190
156 Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys
157                      195                      200                      205
160 Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr
161      210                      215                      220
164 Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
165 225                      230                      235
168 <210> SEQ ID NO: 3
169 <211> LENGTH: 18
170 <212> TYPE: DNA
171 <213> ORGANISM: Artificial Sequence
173 <220> FEATURE:
174 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
175     leader sequence
177 <220> FEATURE:
178 <221> NAME/KEY: CDS
179 <222> LOCATION: (1)..(18)
181 <400> SEQUENCE: 3
182 gca gca gca gca gca gca
183 Ala Ala Ala Ala Ala Ala
184 1      5
187 <210> SEQ ID NO: 4
188 <211> LENGTH: 6
189 <212> TYPE: PRT
190 <213> ORGANISM: Artificial Sequence
192 <220> FEATURE:
193 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
194     leader sequence
196 <400> SEQUENCE: 4
197 Ala Ala Ala Ala Ala Ala
198 1      5
201 <210> SEQ ID NO: 5
202 <211> LENGTH: 18
203 <212> TYPE: DNA
204 <213> ORGANISM: Artificial Sequence
206 <220> FEATURE:
207 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
208     leader sequence
210 <220> FEATURE:
211 <221> NAME/KEY: CDS
212 <222> LOCATION: (1)..(18)
214 <400> SEQUENCE: 5

```

18

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/534,130A

DATE: 05/23/2006

TIME: 14:02:49

Input Set : A:\21415015.APP

Output Set: N:\CRF4\05232006\J534130A.raw

```

215 gcg gcg gcg gcg gcg gcg gcg 18
216 Ala Ala Ala Ala Ala Ala
217 1 5
220 <210> SEQ ID NO: 6
221 <211> LENGTH: 6
222 <212> TYPE: PRT
223 <213> ORGANISM: Artificial Sequence
225 <220> FEATURE:
226 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
227 leader sequence
229 <400> SEQUENCE: 6
230 Ala Ala Ala Ala Ala Ala
231 1 5
234 <210> SEQ ID NO: 7
235 <211> LENGTH: 18
236 <212> TYPE: DNA
237 <213> ORGANISM: Artificial Sequence
239 <220> FEATURE:
240 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
241 leader sequence
243 <220> FEATURE:
244 <221> NAME/KEY: CDS
245 <222> LOCATION: (1)..(18)
247 <400> SEQUENCE: 7
248 gct gct gct gct gct gct 18
249 Ala Ala Ala Ala Ala Ala
250 1 5
253 <210> SEQ ID NO: 8
254 <211> LENGTH: 6
255 <212> TYPE: PRT
256 <213> ORGANISM: Artificial Sequence
258 <220> FEATURE:
259 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
260 leader sequence
262 <400> SEQUENCE: 8
263 Ala Ala Ala Ala Ala Ala
264 1 5
267 <210> SEQ ID NO: 9
268 <211> LENGTH: 18
269 <212> TYPE: DNA
270 <213> ORGANISM: Artificial Sequence
272 <220> FEATURE:
273 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
274 leader sequence
276 <220> FEATURE:
277 <221> NAME/KEY: CDS
278 <222> LOCATION: (1)..(18)
280 <400> SEQUENCE: 9
281 gcc gcc gcc gcc gcc gcc 18

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/534,130A

DATE: 05/23/2006

TIME: 14:02:49

Input Set : A:\21415015.APP

Output Set: N:\CRF4\05232006\J534130A.raw

```

282 Ala Ala Ala Ala Ala Ala
283 1      5
286 <210> SEQ ID NO: 10
287 <211> LENGTH: 6
288 <212> TYPE: PRT
289 <213> ORGANISM: Artificial Sequence
291 <220> FEATURE:
292 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
293     leader sequence
295 <400> SEQUENCE: 10
296 Ala Ala Ala Ala Ala Ala
297 1      5
300 <210> SEQ ID NO: 11
301 <211> LENGTH: 18
302 <212> TYPE: DNA
303 <213> ORGANISM: Artificial Sequence
305 <220> FEATURE:
306 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
307     leader sequence
309 <220> FEATURE:
310 <221> NAME/KEY: CDS
311 <222> LOCATION: (1)..(18)
313 <400> SEQUENCE: 11
314 aga aga aga aga aga aga          18
315 Arg Arg Arg Arg Arg Arg
316 1      5
319 <210> SEQ ID NO: 12
320 <211> LENGTH: 6
321 <212> TYPE: PRT
322 <213> ORGANISM: Artificial Sequence
324 <220> FEATURE:
325 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
326     leader sequence
328 <400> SEQUENCE: 12
329 Arg Arg Arg Arg Arg Arg
330 1      5
333 <210> SEQ ID NO: 13
334 <211> LENGTH: 18
335 <212> TYPE: DNA
336 <213> ORGANISM: Artificial Sequence
338 <220> FEATURE:
339 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
340     leader sequence
342 <220> FEATURE:
343 <221> NAME/KEY: CDS
344 <222> LOCATION: (1)..(18)
346 <400> SEQUENCE: 13
347 cga cga cga cga cga cga          18
348 Arg Arg Arg Arg Arg Arg

```

VERIFICATION SUMMARY

DATE: 05/23/2006

PATENT APPLICATION: US/10/534,130A

TIME: 14:02:50

Input Set : A:\21415015.APP

Output Set: N:\CRF4\05232006\J534130A.raw